Application No.: 10/752,432

REMARKS

Claims 1-34 are pending in this Application of which claims 1, 13, 14, 18, 25, 27, 28, 29 and 34 are independent. Claims 14-28 have been withdrawn from further consideration by the Examiner.

Claims 1-13 and 29-34 have been objected to or rejected as follows:

1.   Claim 34 has been objected to for minor informalities;

2.   Claim 34 has been rejected under 35 U.S.C. § 112, 2nd paragraph, as being indefinite;

3.   Claims 1, 4-13 and 29-33 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900) and further in view of Getchius et al. (U.S. Patent No. 6,493,721);

4.   Claim 2 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Burrows in view of Getchius and further in view of Okabe et al. (U.S. Publication No. 2001/0025287);

5.   Claim 3 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Burrows in view of Getchius and further in view of Zabetian (U.S. Publication No. 2001/0011350); and

6.   Claim 34 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Burrows in view of Dombrowski et al. (U.S. Patent No. 6,233,631).

These rejections are respectfully traversed and are addressed further below.

Claim Objection

The Examiner has objected to the recitation of the term "files extension" in line 4 of claim 34. Claim 34 has been amended as suggested. Withdrawal of the objection is respectfully solicited.

11

Application No.: 10/752,432

Rejection of Claim 34 under 35 U.S.C. § 112

The Examiner indicates that the phrase "that is converted to ASCII text" in lines 4-5 renders claim 34 indefinite because it is unclear whether the phrase refers to "extension" or "document." At least the document is converted to ASCII text. Accordingly, Claim 34 has been amended to clarify this concept.

The Examiner indicates that the phrase "that are indexed for web-based retrieval" in line 7 renders the claim indefinite because it is unclear what is being indexed. The Examiner questions whether the documents, meta-data, text or attachments are indexed. The phrase "identified for retrieval" in line 7 has been deleted because it is repetitious to the phrase that follows. It is believed that this amendment clearly identifies what is being indexed. For example, the claim clearly requires that each document has at least one of either meta-data, text or attachments that are indexed for web-based retrieval.

The Examiner indicates that the phrase "said identification" in line 7 renders the claim indefinite because it is unclear what other element to which this phrase refers. For clarification, claim 34 has been amended to recite that the plurality of documents form a cluster database that is web-searchable.

The Examiner further indicates that the phrase "the cluster database" in line 7 is not found in any of the preceding features of the claim. Accordingly, claim 34 has been amended by removing the term database.

All outstanding indefiniteness issues have been addressed and clarified. Withdrawal of the rejection is respectfully solicited.

Rejection of Claims 1, 4-13 and 29-33 under 35 U.S.C. § 103

The Examiner indicates that claims 1, 4-13 and 29-33 would have been obvious as being unpatentable over Burrows in view of Getchius. Applicants respectfully disagree.

12

Application No.: 10/752,432

Burrows teaches a method for creating an index of records in a database utilizing a search engine. Fig. 2 of Burrows best illustrates a block diagram of the search engine for generating the index. Referring to Fig. 2, the search engine 140 includes an automated web browser 20, a parsing module 30, an indexing module 40, a query module 50, an index stream reader (ISR) 60, an index 70, and a maintenance module 80. During the operation of the search engine 140, the automated browser 20 obtains a number of pages that may be searched. Parsing module 30 breaks down portions of information on the pages 200 into fundamental indexable elements 400, generally parsed by word and its associated location. In addition, the parsing module 30 can detect and encode attributes of the content of the pages. Such attributes may include address 251, description 252, size 253, date 254, fingerprint 255, type 256, and end_page 257. The indexing module 40 sorts word/location pairs 400, which are used to generate the index 70 of the words of the pages 200. During a query, a user interacts with index 70 via query module 50 by providing queries 52. Query request invoke a small number of basic types of object-oriented index stream readers (ISRs) 60, which are used to sequentially scan the data structures 71-73 of the index. This is said to minimize the amount of data that needs to be uncompressed in order to perform a query.

The maintenance module 80 manages the index 70 and its size. In order to reduce size of the index, browser 20 may periodically search the web to determine if a previously indexed page is still active. If the page is no longer active, the browser 20 will inform the maintenance module 80, which will in turn delete associated pages noted in the index 70. See column 14:15-24.

Burrows further teaches that the index size is managed by eliminating duplicates. Specifically, the indexing module 40 identifies duplicate pages by comparing the fingerprint of the current page with the fingerprints of each page identified in the index. In the event that there

13

Application No.: 10/752,432

is a duplicate, the duplicate page (1) will not be identified in the index, or (2) will be updated in the index and the former duplicate page will be deleted from the index. See column 28:48-column 29:5. *It is important to note that duplicate pages are not indexed.*

The Examiner acknowledges that Burrows fails to teach the step of distributing the plurality of native documents and extracted data substantially equally amongst a plurality of nodes of the document management computer system, as claim 1 recites. However, the Examiner alleges that it would have been obvious to combine Burrows with the teachings of Getchius in order to satisfy this claim element.

The Examiner alleges that the combination would have been obvious "so that each node is capable of responding to any search request (column 18, lines 41-43)."

Applicants respectfully disagree.

The combination of references fail to teach each and every element of claim 1. Claim 1 is patentability distinct for several readily apparent reasons.

1.   Claim 1 is directed to "a method for managing a plurality of native documents to be uploaded to a document management computer system." (Emphasis added). At least steps (a) through (e) of claim 1 manage the plurality of native documents. Step (f) also manages the documents by distributing the plurality of native documents and extracted data substantially equally amongst a plurality of nodes of the document management computer system. Conversely, Burrows is directed to the creation of an index of web pages, but not the management and distribution of the web pages.

2.   Burrows fails to disclose step (b) which recites "creating a fingerprint for each native document." Instead, Burrows discloses a parse module 30 that detects page attributes in the meta data for each page. These page attributes can include address, description, size, date,

14

Application No.: 10/752,432

fingerprint, type, and end_page, for example. While the parsing module is able to detect a

fingerprint, there is no disclosure or suggestion that the parsing module creates a fingerprint for

each page, as step (b) of claim 1 would require.

      3.   Burrows fails to disclose the de-duplicating each native document in accordance

with the fingerprint, as claim 1 recites. Burrows does not de-duplicate web page or documents.

Importantly, duplicate web pages remain whereas duplicate index entries are deleted. This is

starkly different from the language of claim 1, which requires that each native document, or in

the case of Burrows each web page, is de-duplicated.

      4.   Burrows fails to disclose or suggest step (f), which recites "distributing the

plurality of native documents and extracted data substantially equally amongst a plurality of

nodes of the document management computer system." However, the Examiner indicates that

Burrows teaches this step except for distributing the information substantially equally.

Applicants respectfully disagree. At least one fundamental difference between Burrows and the

present invention is again noted. Burrows is not concerned with the distribution of the web

pages, but only the creation of an index. There is absolutely no disclosure or suggestion that

Burrows is capable of distributing, or redistributing for that matter, millions of web pages that

make up the internet amongst a plurality of nodes of a document management computer system.

Such would be an unremarkable feat.

      5.   The secondary reference Getchius further fails to disclose the deficiencies

discussed above, especially distributing information substantially equally amongst a plurality of

nodes. The Examiner relies on a statement in the abstract of Getchius which provides that "the

system for performing on line data inquiries is a distributed computer system with a plurality of

server nodes each filly redundant and capable of processing a user query request." The

<div align="center">15</div>

Application No.: 10/752,432

Examiner posits that this statement teaches distributing data substantially equally amongst a

plurality of nodes. Applicants disagree. The Examiner must provide a clear teaching. Getchius

is absent any teaching.

The combination of references fails to teach each and every element of claims 6 and

7.

As regards claim 6, Burrows fails to disclose "discarding native documents that are determined

to be the same in accordance with the comparison of fingerprints," as claim 6 recites.

Specifically, Burrows teaches discarding duplicate indexes. However, the duplicate pages are

not de-duplicated, the index entries are. Burrows fails to disclose or suggest the concepts of

claim 7 for the same reasons.

The combination of references fails to teach each and every element of claims 8-12.

As regards claims 8 and 9, Burrows fails to disclose or suggest creating at least one datafile or a

plurality of datafiles corresponding to the extracted data for each native document. The

Examiner references column 11, line 66-column 12, line 7. However, this text generally teaches

the generation of summary entries in the second level summary data structure whereby each

summary entry includes words associated with the sample. There is absolutely no disclosure or

suggestion that at least one datafile or a plurality of datafiles is created corresponding to the

extracted data for each page 200, as claims 8 and 9 would require. It follows that claims 10 and

11, which depend from claims 8 and 9, respectively, and claim 12 are not taught by Burrows.

The combination of references fails to teach each and every element of claim 13.

The Examiner indicates that the elements of claim 13 are taught for the same reasons as set forth

with respect to claim 1. Likewise, Applicants rely on the remarks above with respect to claim 1.

16

Application No.: 10/752,432

The combination of references fails to teach each and every element of claim 29.

The Examiner refers to the remarks set forth for claim 1. Likewise, Applicants rely on the

remarks set forth herein for claim 1. However, the Examiner posits that Burrows further

discloses a system comprising a computer in connection with the plurality of computer nodes for

receiving a plurality of input files to be uploaded to the plurality of computer nodes. The

Examiner references column 2, lines 51-56 of Burrows which describes a distributed computer

system that includes a database to be indexed. However, again it is noted a fundamental

difference between the present invention and the Burrows patent. Burrows patent is directed to a

management system for indexing a plurality of web pages. The network of Burrows is not

maintained in accordance with the invention described by claim 1. Rather, Burrows is merely

directed to the creation of an index of the web pages on the computer system.

**The combination of references fails to teach each and every element of claims 30, 31,
32, and 33.**

The Examiner refers to the analysis of claims 1, 8 and 10. Likewise, Applicants refer to the

analysis of these claims hereinabove.

Accordingly, the rejection of claims 1, 4-13 and 29-33 have presumably been overcome.

Withdrawal of the rejection is respectfully solicited.

Rejection of claim 2 under 35 U.S.C. § 103

The Examiner indicates that claim 2 is rejected for the reasons set forth for claim 1.

Likewise, Applicants rely on the remarks set forth hereinabove for claim 1. However, the

Examiner further indicates that Burrows and Getchius do not explicitly teach a method

comprising the step of extracting native documents included in the plurality of documents from

an archive file. Applicants agree,

17

WDC99 1040654-1.053217.0014

Application No.: 10/752,432

However, The Examiner indicates that Okabe at page 6, section 0077 teaches the step recited by claim 2. The Examiner indicates that it would have been obvious to make the combination so as to obtain documents from the archive file through the extraction, and that one of ordinary skill in the art would have made the combination with a reasonable expectation of success. Applicants respectfully disagree.

The combination of Burrows, Getchius, and Okabe does not concern a method for managing a plurality of native documents to be uploaded to a document management computer system, as claim 1 recites. Rather, at best, the combination of references is directed to the creation of an index for managing web pages. Even the Okabe reference is not directed to the management of documents existing in a document system. Moreover, there has been no disclosure or suggestion that these references can be applied to a document management computer system in which a plurality of documents are to be uploaded.

The Examiner indicates that the "motivation *obviously* is to obtain documents from the archive through the extraction ...." (*Emphasis Added*) Just because disparate references teach an element of a claim does not necessarily mean that the combination would be especially obvious. The motivation provided by the Examiner is mere hindsight and does not, in any way, establish any motivation for modifying Burrows, which directed to a indexable web server. In fact, Burrows does not disclose the ability to index archived files. Hence, there is no support or rational link for the combination suggested. Accordingly, for the foregoing reasons, the Examiner has failed to establish a *prima facie* case of obviousness of claim 2. Withdrawal of the rejection is respectfully solicited.

Rejection of claim 3 under 35 U.S.C. § 103

The Examiner rejects claim 3 for the same reasons as set forth with respect to claim 1. Likewise, Applicants refer the Examiner to the remarks for claim 1 hereinabove. However, the

18

WDC99 1040654-1.053217.0014

Application No.: 10/752,432

Examiner indicates that the combination of Burrows and Getchius does not explicitly teach a method wherein the fingerprint for each named document is created using a MD5 checksum. Applicants agree. Applicants further emphasize that Burrows and Getchius does not disclose any method for the creation of a fingerprint, as claim 1 recites.

The Examiner relies on Zabetian and posits that it teaches a method wherein the fingerprint for each document is created by using the MD5 checksum. The Examiner refers to page 4, section 0037.

Again, the Examiner is using hindsight reasoning to arrive at the proposed combination. Simply because Zabetian discloses using an MD5 checksum to create a fingerprint does not necessarily mean that the combination of references would have been obvious. In fact, Burrows (or Getchius) fails to teach creating a fingerprint, only detecting a fingerprint. Hence, the no support for the combination suggest. For these reasons, the Examiner has failed to establish a *prima facie* case of obviousness of claim 5. Withdrawal of the rejection is respectfully solicited.

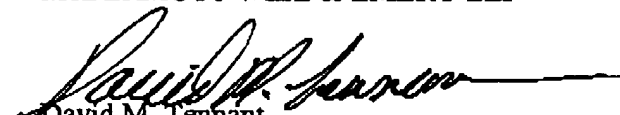### *Rejection of claim 34 under 35 U.S.C. § 103*

Claim 34 has been amended to recite that the electronic documents are stored "substantially equally throughout the cluster," that the operating system generates document fingerprints for each document, and that "said plurality of documents are de-duplicated in accordance with its fingerprint." Accordingly, claim 34 is distinguishable for the same reasons as set forth hereinabove for claim 1. Specifically, Burrows fails to disclose or suggest the newly added claim language. Withdrawal of the rejection is respectfully solicited.

WDC99 1040654-1.053217.0014

Application No.: 10/752,432

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is

hereby made. Please charge any shortage in fees due in connection with the filing of this paper,

including extension of time fees, to Deposit Account 500417 and please credit any excess fees to

such deposit account.

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

David M. Tennant
Registration No. 48,362

600 13th Street, N.W.
Washington, DC 20005-3096
Phone: 202.756.8000 DT/bjs/men/dlb
Facsimile: 202.756.8087
Date: April 21, 2005

**Please recognize our Customer No. 20277
as our correspondence address.**

Certification of Facsimile Transmission

20

WDC99 1040654-1.053217.0014